

Missing values estimation for skylines in incomplete database

ABSTRACT

Incompleteness of data is a common problem in many databases including web heterogeneous databases, multi-relational databases, spatial and temporal databases, and data integration. The incompleteness of data introduces challenges in processing queries as providing accurate results that best meet the query conditions over incomplete database is not a trivial task. Several techniques have been proposed to process queries in incomplete database. Some of these techniques retrieve the query results based on the existing values rather than estimating the missing values. Such techniques are undesirable in many cases as the dimensions with missing values might be the important dimensions of the user's query. Besides, the output is incomplete and might not satisfy the user preferences. In this paper we propose an approach that estimates missing values in skylines to guide users in selecting the most appropriate skylines from the several candidate skylines. The approach utilizes the concept of mining attribute correlations to generate an Approximate Functional Dependencies (AFDs) that captured the relationships between the dimensions. Besides, identify the strength of probability correlations to estimate the values. Then, the skylines with estimated values are ranked. By doing so, we ensure that the retrieved skylines are in the order of their estimated precision.

Keyword: Skyline queries; Preference queries; Incomplete database; Query processing; Estimating missing values.